

## CLAIMS

1. A memory module, comprising:

a plurality of memory devices; and

a memory hub, comprising:

a link interface receiving memory requests for access to memory cells in at least one of the memory devices;

a memory device interface coupled to the memory devices, the memory device interface being operable to couple memory requests to the memory devices for access to memory cells in at least one of the memory devices and to receive read data responsive to at least some of the memory requests;

a history logic unit coupled to the link interface to receive memory requests from the link interface, the history logic being operable to predict addresses that are likely to be accessed in the memory devices based on the memory requests, the history logic unit generating prefetching suggestions indicative of the predicted addresses;

a memory sequencer coupled to the link interface, the memory device interface and the history logic unit, the memory sequencer being operable to couple memory requests to the memory device interface responsive to memory requests received from the link interface, the memory sequencer further being operable to generate and couple prefetch requests to the memory device interface responsive to prefetching suggestions received from the history logic unit;

a prefetch buffer coupled to the memory device interface for receiving and storing read data from memory cells being accessed responsive to the prefetch requests; and

a data read control unit coupled to the memory device interface, the link interface and the prefetch buffer, the data read control circuit being operable to determine from a read memory request received from the link interface if the read data are stored in the prefetch buffer, the control unit further being responsive to the read memory request to couple the read data from the prefetch buffer if the read data are

stored in the prefetch buffer and to couple the read data from the memory device interface if the read data are stored in the memory devices.

2. The memory module of claim 1 wherein the link interface comprises an optical input/output port.

3. The memory module of claim 1 wherein the data read control unit comprises:

a tag logic unit storing prefetch addresses of read data that have been stored in the prefetch buffer responsive to prefetch requests coupled from the memory sequencer to the memory device interface, the tag logic unit further receiving read memory request addresses from the link interface, comparing each read memory request address to the prefetch addresses stored in the tag logic unit, and outputting an active hit signal in the event of an address match and an inactive hit signal in the event of no address match; and

a multiplexer having data inputs coupled to the prefetch buffer and to the memory device interface and a data output coupled to the link interface, the multiplexer being operable to couple the data input from the prefetch buffer to the data output responsive to the active hit signal and to couple the data input from the memory device interface to the data output responsive to the inactive hit signal.

4. The memory module of claim 3 wherein the tag logic unit is further coupled to the memory sequencer to apply the hit signal to the memory sequencer, and wherein the active hit signal prevents the memory sequencer from coupling a memory request to the memory device interface responsive to a memory request received from the link interface.

5. The memory module of claim 1 wherein the memory sequencer is operable to generate and couple prefetch requests to the memory device only when memory requests are not being transferred from the link interface to the memory sequencer.

6. The memory module of claim 1 wherein the history logic unit is further operable to group the predicted addresses into a plurality of sets corresponding to respective strides, and wherein the history logic unit is further operable to cause the prefetch buffer to be divided into a plurality of sections each corresponding to a respective stride, the prefetch buffer storing data read from the memory devices responsive to each of the prefetch requests in the section of the prefetch buffer corresponding to the stride containing the address from which the data was read.

7. The memory module of claim 1 wherein the history logic is operable to selectively enable prefetching based on the nature of the memory requests coupled to the link interface.

8. A memory module, comprising:

a plurality of memory devices; and

a memory hub, comprising:

a link interface receiving memory requests for access to memory cells in at least one of the memory devices;

a memory device interface coupled to the memory devices, the memory device interface being operable to couple memory requests to the memory devices for access to memory cells in at least one of the memory devices and to receive read data responsive to at least some of the memory requests;

a history logic unit coupled to the link interface to receive memory requests from the link interface, the history logic being operable to predict addresses that are likely to be accessed in the memory devices based on the memory requests, the history logic unit generating prefetching suggestions indicative of the predicted addresses;

a memory sequencer coupled to the memory device interface, the memory sequencer being operable to couple prefetch requests to the memory device interface responsive to prefetching suggestions received from the history logic unit;

a prefetch buffer coupled to the memory device interface for receiving and storing read data from memory cells being accessed responsive to the prefetch requests; and

a data read control unit coupled to the memory device interface, the data read control circuit being operable to determine from a read memory request received from the link interface if the read data are stored in the prefetch buffer.

9. The memory module of claim 8 wherein the link interface comprises an optical input/output port.

10. The memory module of claim 8 wherein the data read control unit comprises a tag logic unit storing prefetch addresses of read data that have been stored in the prefetch buffer responsive to prefetch requests coupled from the memory sequencer to the memory device interface, the tag logic unit further receiving read memory request addresses from the link interface, comparing each read memory request address to the prefetch addresses stored in the tag logic unit, and outputting an active hit signal in the event of an address match.

11. The memory module of claim 8 wherein the memory sequencer is operable to generate and couple prefetch requests to the memory device only when memory requests are not being transferred from the link interface to the memory sequencer.

12. The memory module of claim 8 wherein the history logic unit is further operable to group the predicted addresses into a plurality of sets corresponding to respective strides, and wherein the history logic unit is further operable to cause the prefetch buffer to be divided into a plurality of sections each corresponding to a respective stride, the prefetch buffer storing data read from the memory devices responsive to each of the prefetch requests in the section of the prefetch buffer corresponding to the stride containing the address from which the data was read.

13. The memory module of claim 8 wherein the history logic is operable to selectively enable prefetching based on the nature of the memory requests coupled to the link interface.

14. A memory hub, comprising:

- a link interface receiving memory requests;
- a memory device interface operable to output memory requests and to receive read data responsive to at least some of the memory requests;
- a history logic unit coupled to the link interface to receive memory requests from the link interface, the history logic being operable to predict addresses that are likely to be accessed based on the memory requests, the history logic unit generating prefetching suggestions indicative of the predicted addresses;
- a memory sequencer coupled to the link interface, the memory device interface and the history logic unit, the memory sequencer being operable to couple memory requests to the memory device interface responsive to memory requests received from the link interface, the memory sequencer further being operable to generate and couple prefetch requests to the memory device interface responsive to prefetching suggestions received from the history logic unit;
- a prefetch buffer coupled to the memory device interface for receiving and storing read data received responsive to the prefetch requests; and
- a data read control unit coupled to the memory device interface, the link interface and the prefetch buffer, the data read control circuit being operable to determine from a read memory request received from the link interface if the read data are stored in the prefetch buffer, the control unit further being responsive to the read memory request to couple the read data from the prefetch buffer if the read data are stored in the prefetch buffer and to couple the read data from the memory device interface if the read data are not stored in the prefetch buffer.

15. The memory hub of claim 14 wherein the link interface comprises an optical input/output port.

16. The memory hub of claim 14 wherein the data read control unit comprises:

a tag logic unit storing prefetch addresses of read data that have been stored in the prefetch buffer responsive to prefetch requests coupled from the memory sequencer to the memory device interface, the tag logic unit further receiving read memory request addresses from the link interface, comparing each read memory request address to the prefetch addresses stored in the tag logic unit, and outputting an active hit signal in the event of an address match and an inactive hit signal in the event of no address match; and

a multiplexer having data inputs coupled to the prefetch buffer and to the memory device interface and a data output coupled to the link interface, the multiplexer being operable to couple the data input from the prefetch buffer to the data output responsive to the active hit signal and to couple the data input from the memory device interface to the data output responsive to the inactive hit signal.

17. The memory hub of claim 16 wherein the tag logic unit is further coupled to the memory sequencer to apply the hit signal to the memory sequencer, and wherein the active hit signal prevents the memory sequencer from coupling a memory request to the memory device interface responsive to a memory request received from the link interface.

18. The memory hub of claim 14 wherein the memory sequencer is operable to generate and couple prefetch requests to the memory device only when memory requests are not being transferred from the link interface to the memory sequencer.

19. The memory hub of claim 14 wherein the history logic unit is further operable to group the predicted addresses into a plurality of sets corresponding to respective strides, and wherein the history logic unit is further operable to cause the prefetch buffer to be divided into a plurality of sections each corresponding to a respective stride, the prefetch buffer storing data read from the memory devices responsive to each of the prefetch requests

in the section of the prefetch buffer corresponding to the stride containing the address from which the data was read.

20. The memory hub of claim 14 wherein the history logic is operable to selectively enable prefetching based on the nature of the memory requests coupled to the link interface.

21. A memory hub, comprising:

- a link interface receiving memory requests;
- a memory device interface operable to output memory requests and to receive read data responsive to at least some of the memory requests;
- a history logic unit coupled to the link interface to receive memory requests from the link interface, the history logic being operable to predict addresses that are likely to be accessed based on the memory requests, the history logic unit generating prefetching suggestions indicative of the predicted addresses;
- a memory sequencer coupled to the memory device interface and the history logic unit, the memory sequencer being operable to couple prefetch requests to the memory device interface responsive to prefetching suggestions received from the history logic unit;
- a prefetch buffer coupled to the memory device interface for receiving and storing read data accessed responsive to the prefetch requests; and
- a data read control unit coupled to the memory device interface and the prefetch buffer, the data read control circuit being operable to determine from a read memory request received from the link interface if the read data are stored in the prefetch buffer and to couple the read data from the prefetch buffer responsive to determining that the read data are stored in the prefetch buffer.

22. The memory hub of claim 21 wherein the link interface comprises an optical input/output port.

23. The memory hub of claim 21 wherein the data read control unit comprises a tag logic unit storing prefetch addresses of read data that have been stored in the prefetch buffer responsive to prefetch requests coupled from the memory sequencer to the memory device interface, the tag logic unit further receiving read memory request addresses from the link interface, comparing each read memory request address to the prefetch addresses stored in the tag logic unit, and outputting an active hit signal in the event of an address match.

24. The memory hub of claim 21 wherein the memory sequencer is operable to generate and couple prefetch requests to the memory device only when memory requests are not being transferred from the link interface to the memory sequencer.

25. The memory hub of claim 21 wherein the history logic unit is further operable to group the predicted addresses into a plurality of sets corresponding to respective strides, and wherein the history logic unit is further operable to cause the prefetch buffer to be divided into a plurality of sections each corresponding to a respective stride, the prefetch buffer storing data read from the memory devices responsive to each of the prefetch requests in the section of the prefetch buffer corresponding to the stride containing the address from which the data was read.

26. The memory hub of claim 21 wherein the history logic is operable to selectively enable prefetching based on the nature of the memory requests coupled to the link interface.

27. A computer system, comprising:  
a central processing unit ("CPU");  
a system controller coupled to the CPU, the system controller having an input port and an output port;  
an input device coupled to the CPU through the system controller;  
an output device coupled to the CPU through the system controller;

a storage device coupled to the CPU through the system controller;  
a plurality of memory modules, each of the memory modules comprising:

a plurality of memory devices; and

a memory hub, comprising:

a link interface receiving memory requests for access to memory cells in at least one of the memory devices;

a memory device interface coupled to the memory devices, the memory device interface being operable to couple memory requests to the memory devices for access to memory cells in at least one of the memory devices and to receive read data responsive to at least some of the memory requests;

a history logic unit coupled to the link interface to receive memory requests from the link interface, the history logic being operable to predict addresses that are likely to be accessed in the memory devices based on the memory requests, the history logic unit generating prefetching suggestions indicative of the predicted addresses;

a memory sequencer coupled to the link interface, the memory device interface and the history logic unit, the memory sequencer being operable to couple memory requests to the memory device interface responsive to memory requests received from the link interface, the memory sequencer further being operable to generate and couple prefetch requests to the memory device interface responsive to prefetching suggestions received from the history logic unit;

a prefetch buffer coupled to the memory device interface for receiving and storing read data from memory cells being accessed responsive to the prefetch requests; and

a data read control unit coupled to the memory device interface, the link interface and the prefetch buffer, the data read control circuit being operable to determine from a read memory request received

from the link interface if the read data are stored in the prefetch buffer, the control unit further being responsive to the read memory request to couple the read data from the prefetch buffer if the read data are stored in the prefetch buffer and to couple the read data from the memory device interface if the read data are stored in the memory devices.

28. The computer system of claim 27 wherein the link interface comprises an optical input/output port.

29. The computer system of claim 27 wherein the data read control unit comprises:

a tag logic unit storing prefetch addresses of read data that have been stored in the prefetch buffer responsive to prefetch requests coupled from the memory sequencer to the memory device interface, the tag logic unit further receiving read memory request addresses from the link interface, comparing each read memory request address to the prefetch addresses stored in the tag logic unit, and outputting an active hit signal in the event of an address match and an inactive hit signal in the event of no address match; and

a multiplexer having data inputs coupled to the prefetch buffer and to the memory device interface and a data output coupled to the link interface, the multiplexer being operable to couple the data input from the prefetch buffer to the data output responsive to the active hit signal and to couple the data input from the memory device interface to the data output responsive to the inactive hit signal.

30. The computer system of claim 29 wherein the tag logic unit is further coupled to the memory sequencer to apply the hit signal to the memory sequencer, and wherein the active hit signal prevents the memory sequencer from coupling a memory request to the memory device interface responsive to a memory request received from the link interface.

31. The computer system of claim 27 wherein the memory sequencer is operable to generate and couple prefetch requests to the memory device only when memory requests are not being transferred from the link interface to the memory sequencer.

32. The computer system of claim 27 wherein the history logic unit is further operable to group the predicted addresses into a plurality of sets corresponding to respective strides, and wherein the history logic unit is further operable to cause the prefetch buffer to be divided into a plurality of sections each corresponding to a respective stride, the prefetch buffer storing data read from the memory devices responsive to each of the prefetch requests in the section of the prefetch buffer corresponding to the stride containing the address from which the data was read.

33. The computer system of claim 27 wherein the history logic is operable to selectively enable prefetching based on the nature of the memory requests coupled to the link interface.

34. A computer system, comprising:

- a central processing unit ("CPU");
- a system controller coupled to the CPU, the system controller having an input port and an output port;
- an input device coupled to the CPU through the system controller;
- an output device coupled to the CPU through the system controller;
- a storage device coupled to the CPU through the system controller;
- a plurality of memory modules, each of the memory modules comprising:
  - a plurality of memory devices; and
  - a memory hub, comprising:
    - a link interface receiving memory requests for access to memory cells in at least one of the memory devices;
    - a memory device interface coupled to the memory devices, the memory device interface being operable to couple memory requests to

the memory devices for access to memory cells in at least one of the memory devices and to receive read data responsive to at least some of the memory requests;

a history logic unit coupled to the link interface to receive memory requests from the link interface, the history logic being operable to predict addresses that are likely to be accessed in the memory devices based on the memory requests, the history logic unit generating prefetching suggestions indicative of the predicted addresses;

a memory sequencer coupled to the memory device interface and the history logic unit, the memory sequencer being operable to couple prefetch requests to the memory device interface responsive to prefetching suggestions received from the history logic unit;

a prefetch buffer coupled to the memory device interface for receiving and storing read data from memory cells being accessed responsive to the prefetch requests; and

a data read control unit coupled to the memory device interface and the prefetch buffer, the data read control circuit being operable to determine from a read memory request received from the link interface if the read data are stored in the prefetch buffer and to couple the read data from the prefetch buffer responsive to determining that the read data are stored in the prefetch buffer.

35. The computer system of claim 34 wherein the link interface comprises an optical input/output port.

36. The computer system of claim 34 wherein the data read control unit comprises a tag logic unit storing prefetch addresses of read data that have been stored in the prefetch buffer responsive to prefetch requests coupled from the memory sequencer to the memory device interface, the tag logic unit further receiving read memory request addresses

from the link interface, comparing each read memory request address to the prefetch addresses stored in the tag logic unit, and outputting an active hit signal in the event of an address match.

37. The computer system of claim 34 wherein the memory sequencer is operable to generate and couple prefetch requests to the memory device only when memory requests are not being transferred from the link interface to the memory sequencer.

38. The computer system of claim 34 wherein the history logic unit is further operable to group the predicted addresses into a plurality of sets corresponding to respective strides, and wherein the history logic unit is further operable to cause the prefetch buffer to be divided into a plurality of sections each corresponding to a respective stride, the prefetch buffer storing data read from the memory devices responsive to each of the prefetch requests in the section of the prefetch buffer corresponding to the stride containing the address from which the data was read.

39. The computer system of claim 34 wherein the history logic is operable to selectively enable prefetching based on the nature of the memory requests coupled to the link interface.

40. A method of reading data from a memory module, comprising:

- receiving memory requests for access to a memory device mounted on the memory module;
- coupling the memory requests to the memory device responsive to the received memory request, at least some of the memory requests being memory requests to read data;
- receiving read data responsive to the read memory requests;
- predicting addresses that are likely to be accessed in the memory device based on the read memory requests, the address prediction being internal to the memory module;
- generating prefetching suggestions indicative of the predicted addresses;
- generating prefetch requests responsive to the prefetching suggestions;

coupling the prefetch requests to the memory device;  
receiving prefetched read data responsive to the prefetch requests;  
storing the prefetched read data in a prefetch buffer;  
determining from a read memory request if the requested read data are stored  
in the prefetch buffer;

coupling the read data from the prefetch buffer if a determination has been  
made that the read data are stored in the prefetch buffer; and

coupling the read data from the memory device if a determination has not been  
made that the read data are stored in the prefetch buffer.

41. The method of claim 40 wherein the act of receiving memory requests  
for access to a memory device mounted on the memory module comprises receiving optical  
signals corresponding to the memory requests.

42. The method of claim 40 wherein the act of determining from a read  
memory request if the requested read data are stored in the prefetch buffer comprises:

storing prefetch addresses of read data that have been stored in the prefetch  
buffer responsive to the prefetch requests;

receiving read memory request addresses;

comparing each received read memory request address to the prefetch  
addresses stored in the tag logic unit; and

determining that the requested read data are stored in the prefetch buffer in the  
event of an address match.

43. The method of claim 42, further comprising:

coupling read data from the prefetch buffer responsive to determining that the  
requested read data are stored in the prefetch buffer; and

coupling read data from the memory device responsive to determining that the  
requested read data are not stored in the prefetch buffer.

44. The method of claim 40 wherein the act of coupling the prefetch requests to the memory device comprises coupling the prefetch requests to the memory device only when the memory requests are not being received.

45. The method of claim 40, further comprising selectively enabling prefetching based on the nature of the received memory requests.

46. The method of claim 40, further comprising:

- grouping the predicted addresses into a plurality of sets corresponding to respective strides;
- dividing the prefetch buffer into a plurality of sections each corresponding to a respective stride; and
- storing data read from the memory device responsive to each of the prefetch requests in the section of the prefetch buffer corresponding to the stride containing the address from which the data was read.

47. A method of reading data from a memory module, comprising:

- receiving memory requests for access to a memory device mounted on the memory module;
- coupling the memory requests to the memory device responsive to the received memory request, at least some of the memory requests being memory requests to read data;
- receiving read data responsive to the read memory requests;
- predicting addresses that are likely to be accessed in the memory device based on the read memory requests, the address prediction being internal to the memory module;
- generating prefetching suggestions indicative of the predicted addresses;
- generating prefetch requests responsive to the prefetching suggestions;
- coupling the prefetch requests to the memory device;
- receiving prefetched read data responsive to the prefetch requests;
- storing the prefetched read data in a prefetch buffer;

determining from a read memory request if the requested read data are stored in the prefetch buffer; and

coupling the read data from the prefetch buffer if a determination has been made that the read data are stored in the prefetch buffer.

48. The method of claim 47 wherein the act of receiving memory requests for access to a memory device mounted on the memory module comprises receiving optical signals corresponding to the memory requests.

49. The method of claim 47 wherein the act of determining from a read memory request if the requested read data are stored in the prefetch buffer comprises:

storing prefetch addresses of read data that have been stored in the prefetch buffer responsive to the prefetch requests;

receiving read memory request addresses;

comparing each received read memory request address to the prefetch addresses stored in the tag logic unit; and

determining that the requested read data are stored in the prefetch buffer in the event of an address match.

50. The method of claim 49, further comprising:

coupling read data from the prefetch buffer responsive to determining that the requested read data are stored in the prefetch buffer; and

coupling read data from the memory device responsive to determining that the requested read data are not stored in the prefetch buffer.

51. The method of claim 47 wherein the act of coupling the prefetch requests to the memory device comprises coupling the prefetch requests to the memory device only when the memory requests are not being received.

52. The method of claim 47, further comprising:

grouping the predicted addresses into a plurality of sets corresponding to respective strides;

dividing the prefetch buffer into a plurality of sections each corresponding to a respective stride; and

storing data read from the memory device responsive to each of the prefetch requests in the section of the prefetch buffer corresponding to the stride containing the address from which the data was read.

53. The method of claim 47, further comprising selectively enabling prefetching based on the nature of the received memory requests.